

SAM-IE: SAM-based image enhancement for facilitating medical image diagnosis with segmentation foundation model

Changyan Wang^{a,b}, Haobo Chen^{a,b}, Xin Zhou^{a,b}, Meng Wang^{a,b}, Qi Zhang^{a,b,*}

^a School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

^b The SMART (Smart Medicine and AI-based Radiology Technology) Lab, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Keywords:

Segment Anything Model
Medical image classification
Image enhancement

ABSTRACT

The Segment Anything Model (SAM) is a large-scale model developed for general segmentation tasks in computer vision. Trained on a substantial dataset, SAM can accurately segment various objects in natural scene images. However, due to significant semantic differences between medical and natural images, directly applying SAM to medical image segmentation does not yield optimal results. Therefore, effectively utilizing such a comprehensive foundation model for medical image analysis is an emerging research topic. Despite SAM's current suboptimal performance in medical image segmentation, it shows preliminary recognition and localization of tissues and lesions that radiologists focus on in medical images. This implies that SAM's generated masks, features, and stability scores hold potential value for medical image diagnosis. Therefore, based on the model output of SAM, this study introduces a SAM-based Image Enhancement (SAM-IE) method for disease diagnosis. Targeting popular medical image classification models (e.g., ResNet50 and Swin Transformer), SAM-IE is proposed to enhance image inputs by combining the binary mask and contour mask generated by SAM with the original image to create attention maps, thereby improving diagnostic performance. To validate the effectiveness of SAM-IE for diagnosis, experiments were conducted on four medical image datasets for eight classification tasks. The results demonstrate the effectiveness of our proposed SAM-IE model, showcasing SAM's potential value in medical image classification. This study provides a feasible approach for integrating SAM into disease diagnosis.

1. Introduction

The Segment Anything Model (SAM) emerges as an innovative foundational model for image segmentation in 2023, leveraging the vision transformer architecture (Dosovitskiy et al., 2021; Kirillov et al., 2023). The SAM comprises a vision transformer-based image encoder, a prompt encoder, and a lightweight mask decoder. The image features extracted by SAM's encoder are used by the mask decoder to generate segmentation results, incorporating the embedded prompt information. Trained on an extensive dataset comprising 11 million images with 1 billion masks, SAM stands out for its notable zero-shot segmentation performance on previously unseen datasets and tasks (Huang et al., 2023; Mazurowski et al., 2023). One of SAM's key strengths lies in its versatility, demonstrating efficacy across a diverse range of segmentation tasks.

While SAM demonstrates impressive performance in natural image

segmentation, studies indicate that it may face limitations in segmentation tasks requiring domain-specific knowledge, as observed in certain medical image segmentation scenarios (Mazurowski et al., 2023; Zhang & Jiao, 2023). Deng et al. (2023) assessed SAM's performance in tumor segmentation, non-tumor tissue segmentation, and cell nuclei segmentation. Even with 20 prompts on each image, SAM failed to achieve satisfactory performance for dense instance object segmentation. Hu, Xia, Ju, and Li (2023) performed experiments on multi-phase liver tumor segmentation using contrast-enhanced computed tomography volumes. The results indicated a significant gap between SAM with a limited number of prompt points and the classic U-Net (Ronneberger, Fischer, & Brox, 2015). Zhou et al. (2023) assessed the performance of SAM in segmenting polyps from colonoscopy images across five benchmark datasets in an unprompted setting. The experimental results revealed SAM's lower performance compared to state-of-the-art methods.

* Corresponding author at: School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China.

E-mail addresses: wangcy@shu.edu.cn (C. Wang), haobochen@shu.edu.cn (H. Chen), zhouxinsya@shu.edu.cn (X. Zhou), wangmeng0102@shu.edu.cn (M. Wang), zhangq@t.shu.edu.cn (Q. Zhang).

<https://doi.org/10.1016/j.eswa.2024.123795>

Received 22 December 2023; Received in revised form 21 February 2024; Accepted 18 March 2024

Available online 22 March 2024

0957-4174/© 2024 Elsevier Ltd. All rights reserved.

Many research findings highlight limitations in the segmentation capability of SAM when directly applied to medical image, revealing significant discrepancies across various datasets and tasks. While SAM demonstrates exceptional performance in specific objects and modalities, it falls short and may even fail in more challenging scenarios, particularly when dealing with weak boundaries, low contrast, and small, irregular shapes (G.-P. Ji et al., 2023; W. Ji et al., 2023). The challenges faced by SAM in medical image segmentation are attributed to the intricate modalities, fine anatomical structures, uncertain and complex object boundaries, and a wide range of object scales (Huang et al., 2023). Furthermore, being pre-trained on the SA-1B dataset, which contains 11 million natural images, SAM's approach of determining boundaries based on intensity variance (G.-P. Ji et al., 2023; Ma et al., 2023; Zhang & Jiao, 2023), effective in natural images, proves inadequate for medical images due to the crucial role of anatomical or pathological structure analysis. Additionally, SAM struggles to associate segmentation regions with meaningful semantic classes, hindering its application in computer-aided diagnosis.

In recent years, many studies have been dedicated to improving SAM to enhance its applicability in medical image analysis. Due to the less-than-ideal results of directly applying SAM to medical image segmentation, several studies have focused on how to fine-tune SAM for medical image analysis to enhance its reliability. Hu, Li, and Yang (2023) and Li, Hu, and Yang (2023) conducted fine-tuning experiments on SAM for skin cancer and polyp segmentation tasks, respectively, achieving promising segmentation results. While fine-tuning SAM on medical datasets holds the potential to improve its segmentation performance, its efficacy heavily relies on manually provided prompt information and is sensitive to incorrect prompts. In addressing this issue, Gao et al. (Gao, Xia, Hu, & Gao, 2023) proposed the Decoupled Segment Anything Model, aiming to minimize performance degradation caused by erroneous prompts. These studies attempt to restructure SAM to adapt it for medical image segmentation tasks and enhance the level of segmentation in medical images.

Due to the model being trained on a large number of natural images, SAM has significant limitations in the field of medical image segmentation. Currently, research based on SAM in the field of medical image analysis mostly focuses on improving the segmentation performance on medical images. However, existing studies indicate that it appears to be difficult for SAM to achieve high-quality segmentation of medical images, whether through model fine-tuning or adding prompts. Although SAM cannot accurately segment medical images without prompting, SAM can still distinguish obvious tissues and pathological structures according to pixel changes in the images. Medical images carry a substantial amount of information, primarily associated with high image resolution and pixel depth, which can exceed the visual discernment capabilities of the human eye in distinguishing among numerous gray levels (Ortiz, Górriz, Ramírez, Salas-González, & Llamas-Elvira, 2013). Enhancing the appearance and visual quality of medical images is crucial to provide physicians with valuable information that may not be immediately evident in the original image. This enhancement assists in anomaly detection, diagnosis, and treatment. In this context, the image enhancement (IE) techniques aim to achieve specific improvements in the quality of a given medical image. The enhanced image is expected to better reveal certain features compared to their original appearance (de Araujo, Constantinou, & Tavares, 2014). Therefore, in this paper we propose a new IE method based on SAM (SAM-IE), aiming at improving the diagnostic accuracy of medical image classification models. Our aim is to explore SAM's potential value in medical image analysis from a different perspective. SAM-generated masks and stability scores, without additional prompts, prove useful for medical image classification and diagnosis. Thus, we introduce SAM-IE to enhance inputs for medical image classification models.

A critical difference between SAM-IE and the previous enhancement methods (Dinh & Giang, 2022; Rundo et al., 2019) is that the traditional IE methods often work at a low level, e.g., de-blurring and noise

reduction, and the purpose of enhancement is for image reconstruction and recovery. In contrast, SAM-IE aims to add high-level structures to original images, providing better semantics for the subsequent medical image classification models. The SAM-IE enhances images by adding semantic structures from a segmentation foundation model. Moreover, the SAM-IE opens up new research ideas for SAM in the field of medical image analysis and further explores the application values of SAM. On the other hand, the image enhancement method proposed in this study does not involve complex modifications to the SAM or excessive prior prompts, making it convenient for radiologists to directly use SAM in the disease diagnosis process.

In order to test the effect of SAM-IE integrated into medical image classification models for disease diagnosis, this paper selects two classification models commonly used in medical image classification tasks, and carries out classification experiments on four medical image datasets respectively. The effect of SAM-IE on medical image classification task was analyzed by comparing the classification results before and after SAM-IE was used on classification models. Our main contributions can be summarized as follows: (1) We propose an image enhancement method based on the Segment Anything Model. (2) The SAM-IE method we propose can enhance the performance of classification models. (3) The SAM-IE demonstrates effectiveness for disease diagnosis across various imaging modalities. (4) Our method extends the scope of SAM to medical image diagnosis. The second chapter reviews the research on application of SAM in medical image analysis. In the third chapter of this paper, the SAM-IE methods and model training and deployment processes are explained in detail. The datasets, the details of the experiment and the evaluation indicators are described in chapter four. The fifth part of the paper mainly analyzes the experimental results on four datasets.

2. Related work

2.1. Optimization of SAM on medical images

The research shows that it is difficult to obtain satisfactory segmentation effect by directly applying SAM to medical image segmentation task. In order to make better use of SAM in the field of medical image analysis, many researches focus on the transformation of SAM and the integration of SAM with other methods. Among them, the method of fine-tuning a small part of SAM parameters to obtain better medical image segmentation has attracted more attention. Ma et al. (2023) introduce MedSAM for universal image segmentation by curating a diverse and comprehensive medical image dataset containing over 200,000 masks with 11 modalities and develop fine-tuning approach to adapt SAM to medical image segmentation. The proposed MedSAM further improves the performance of SAM in medical image segmentation. Wu et al. (2023) introduce Medical SAM Adapter to fine-tuning pre-trained SAM with a parameter-efficient fine-tuning paradigm using Adaption modules (Hu, 2021). Comprehensive experiments demonstrate that by fine-tuning, Medical SAM Adapter can obtain comparable performance compared with state-of-the-art methods. Zhang and Liu (2023) apply low-rank-based (Hu, 2021) fine-tuning strategy to SAM image encoder together with the prompt encoder and mask decoder on labeled medical image segmentation datasets. By fine-tuning on a multi-organ segmentation dataset, SAM can achieve highly competitive segmentation performance compared with state-of-the-art methods. Chai et al. (2023) combine an additional CNN as a complementary encoder along with the standard SAM architecture and only focus on fine-tuning the additional CNN and SAM decoder to reduce the resource utilization and training time of fine-tuning.

As simple and straight-forward approaches, these methods demonstrate the effectiveness of fine-tuning SAM on domain-specific medical datasets to achieve better segmentation performance. However, the use of SAM for medical image segmentation still needs to provide additional professional tips, and it is difficult to achieve fully automatic medical

image segmentation. For the prompt mode, the final segmentation results are highly dependent on the prompt, and the model is still more sensitive to error prompts. To issue this challenge, Gao et al. (2023) propose the Decoupling Segment Anything Model, which can minimize the performance degradation caused by wrong prompts while avoiding training image encoder which requires higher GPU cost. Extensive experiments demonstrate that the Decoupling Segment Anything Model improves the robustness of fully automated segmentation in dealing with distribution variations across different sites. Deng, et al. (2023) propose to enhance SAM by employing multiple box prompts to establish pixel-level reliability through uncertainty estimation. By generating different predictions using different multi-box prompts and estimating the distribution of SAM predictions using Monte Carlo simulation with prior distribution parameters, the model can estimate the variations by aleatoric uncertainty and generate an uncertainty map to highlight challenging areas for segmentation, which offers valuable guidance for potential segmentation errors and support further clinical analysis.

Direct application of SAM trained on a large number of natural images to medical image segmentation often fails to achieve satisfactory segmentation results. While methods for fine-tuning SAM can enhance its performance to some extent in medical image segmentation, specific domain expertise is still required for medical image segmentation tasks, making fully automatic segmentation tasks difficult to achieve. Other methods for improving SAM inevitably require higher GPU costs and training dataset costs. Therefore, this paper proposes an image enhancement method based on SAM. This method no longer blindly pursues higher performance of SAM in medical image segmentation but fully utilizes the masks and stability scores generated by SAM to enhance medical images, aiding physicians in disease diagnosis. At the same time, our method further expands the application scope of SAM in the field of medical image diagnosis.

2.2. Usability extension of SAM on medical images

In contrast to natural images, medical images are typically stored in specific formats like NII and DICOM. To facilitate SAM's integration into medical image analysis, Liu, Zhang, She, Kheradmand and Mehran (2023) incorporated SAM into the 3D Slicer software (Fedorov et al., 2012). This integration allows researchers to perform segmentation tasks on medical images with minimal latency, and the segmentation process, initiated by a prompt, automatically propagates to the subsequent slice once the segmentation for a particular slice is completed (Liu et al., 2023; Zhang & Jiao, 2023).

To improve the performance of interactive image segmentation, Dai et al. (2023) introduced SAMAug, an innovative visual point augmentation technique designed for SAM. SAMAug creates augmented point prompts to convey additional information about the user's intent to SAM. Starting with an initial point prompt, SAM generates an initial mask, further processed by SAMAug to generate augmented point prompts. Integrating these additional points enables SAM to produce augmented segmentation masks, leading to enhanced segmentation performance.

While SAM's performance depends on input prompts, there is a growing interest in achieving a fully automatic solution. Shaharabany, Dahan, Giryes and Wolf (2023) introduced AutoSAM, incorporating the training of an auxiliary prompt encoder to generate a surrogate prompt. Unlike traditional prompt types in SAM, such as bounding boxes, points, or masks, AutoSAM employs the image itself as input. With the assistance of the auxiliary trained network, SAM transitions into a fully automatic mode, eliminating the need for prompts and achieving state-of-the-art results across various medical benchmarks without fine-tuning.

Cui et al. (2023) presented an approach called all-in-SAM, designed to leverage SAM without manual prompts. This framework specifically exploits weak annotations and pre-trained SAM for fine-tuning, aiming to minimize annotation costs and enhance SAM's application through

label-efficient fine-tuning. Additionally, Lei, Xu, Zhang, Kang and Zhang (2023) proposed MedLSAM, introducing a localization process by identifying six extreme points in three directions of 3D images for any region of interest. Subsequently, the generated bounding box is utilized by SAM for precise segmentation of the target anatomy, enabling automatic segmentation.

Currently, the application of SAM in the field of medical image analysis mostly focuses on image segmentation tasks. Therefore, the segmentation performance of SAM on medical images limits further utilization of SAM in medical image analysis. This paper proposes an image enhancement method based on SAM, challenging the current view of SAM as only a segmentation tool and presenting a novel approach for utilizing SAM as an image enhancement tool in medical image diagnosis.

2.3. Medical image augmentation with SAM

Different from previous studies, which dedicate to improving the medical image segmentation effect of SAM. Zhang, Zhou, Wang, Liang, and Chen (2023) did not directly apply SAM for segmentation, but utilized the segmentation masks generated by SAM to augment the original input medical images. Their proposed SAMAug fuses the original image with the segmented prior image and the boundary prior image generated by the segmented prior image. The experimental results show that the segmentation effect after input augmentation with SAMAug is better than that without augmentation. This study demonstrates that the SAM may not be able to generate high-quality medical image segmentation, but these generated masks and features still help boost the segmentation model.

Research indicates that combining SAM's output images with original images to generate prior maps can be used to enhance network inputs, thereby improving the performance of downstream medical segmentation models. Inspired by this, this paper fully utilizes the masks and stability scores outputted by SAM to design the SAM-IE image enhancement model. The enhanced images can effectively highlight the pathological regions in the original medical images and provide attention maps for medical image classification models, thereby enhancing the classification performance of the models. The SAM-IE method proposed in this paper demonstrates effectiveness in disease diagnosis across multiple medical image modalities.

3. Methods

The pre-trained model of SAM currently does not achieve the same level of segmentation performance in medical images as it does in natural images. However, SAM still excels in highlighting certain noteworthy lesion areas or typical features in medical images. From this perspective, SAM is expected to assist in disease diagnosis by enhancing image for the classification model. In order to study the impact of SAM-IE on the classification performance of medical images, this study uses the common medical image classification models (e.g., ResNet50 (Kaiming, Xiangyu, Shaoqing, & Jian, 2016) and Swin Transformer (Liu et al., 2021)) to conduct classification experiments on the pre-enhanced and post-enhanced images respectively, and analyzes and compares the two groups of classification results, as shown in Fig. 1. In Section 3.1, we describe the process of generating binary mask and contour mask by SAM, and elaborates the process of IE by applying two key masks to medical images. Section 3.2 will introduce the training details of the medical image classification model using SAM-IE, and explain the model testing methods. Section 3.3 will primarily detail the steps for evaluating the performance of the classification networks with SAM-IE.

3.1. Image enhancement with SAM

Loading pre-trained SAM allows for the generation of segmentation masks for medical images. Without adding any prior prompts, SAM is

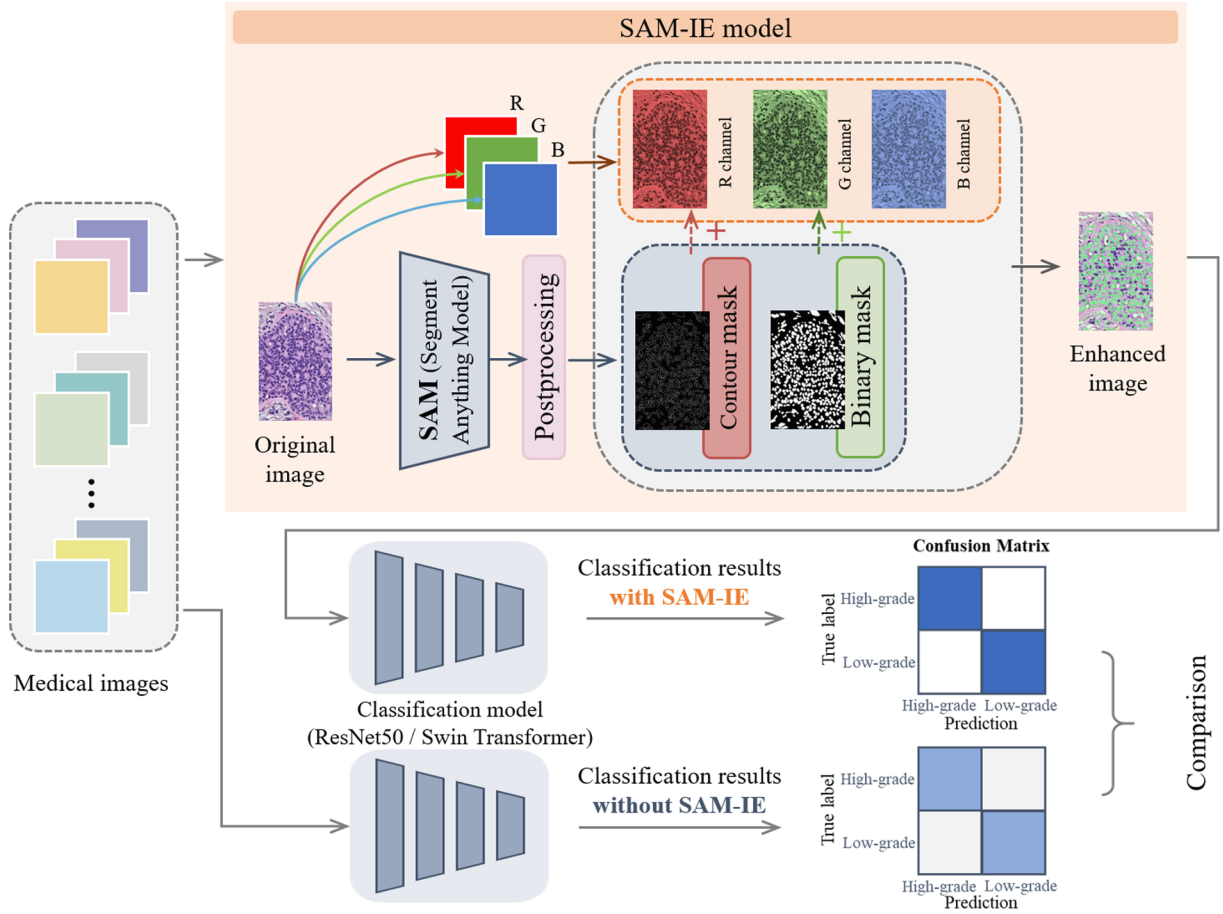


Fig. 1. The flowchart of the medical image classification with SAM-based Image Enhancement (SAM-IE). The terms ‘low-grade’ and ‘high-grade’ can refer to benign and malignant, respectively, or to different degrees of disease severity.

capable of generating segmentation masks for all possible regions in a medical image and stores them in a list. Moreover, in SAM segmentation experiments, stability scores corresponding to each segmentation mask are also outputted along with the segmentation masks. In this paper, based on the characteristics of different medical images, segmentation masks are filtered according to the stability scores. The stability scores of the retained segmentation masks are mapped to grayscale values and used to generate a binary mask, which is a grayscale image depicting the corresponding regions. Additionally, this paper further extracts the edge information from all the filtered segmentation masks to generate a contour mask. In summary, for a given medical image, this paper completes the drawing of the medical image binary mask and contour mask through the above process, as seen in Fig. 1.

After obtaining the binary mask and contour mask generated by SAM, this study enhances the images by overlaying these two masks with the original medical image. Many medical image segmentation tasks can be simplified into three types of segmentation tasks, where the first type corresponds to the background, the second type corresponds to the region of interest (ROI), and the third type corresponds to the boundary between ROI and background. Therefore, for three-channel color images, we generate enhanced images by overlaying binary masks and contour masks with a single channel of the original image. Specifically, we first split the original medical image into R, G, and B channels. Then, we overlay the contour mask onto the R channel. Similarly, we overlay the binary mask onto the G channel. Finally, we combine the newly generated R and G channels with the original B channel to obtain the final enhanced image (Fig. 1). For medical images where the original image is grayscale, we create a three-channel image where the first channel consists of the grayscale original image, the

second channel consists of the overlay of the original grayscale image with the binary mask, and the third channel consists of the overlay of the original grayscale image with the contour mask. For each medical image x in the training set, its enhanced version can be represented as $x^{IE} = IE(Mask_{contour}, Mask_{binary}, x)$.

3.2. Classification model training and testing with SAM-IE

The original training set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where $x_i \in \mathbb{R}^{w \times h \times 3}$, $y_i \in \{0, 1\}$ is the classification label of the medical image x_i . By applying SAM-IE to each medical image in the training set, a new enhanced training set is generated $\{(x_1^{IE}, y_1), (x_2^{IE}, y_2), \dots, (x_n^{IE}, y_n)\}$, where $x_i^{IE} \in \mathbb{R}^{w \times h \times 3}$ is the enhanced version of the medical image x_i . For the training set that was not enhanced by SAM-IE, this paper employs a common medical image classification model M (e.g., ResNet50 and Swin Transformer) to learn from it. The parameters of M are optimized through the following learning objectives:

$$\sum_{i=1}^n \text{loss}(M(x_i), y_i) \quad (1)$$

As for the images enhanced by SAM-IE, this paper trains the classification model M in the same way as before enhancement. However, considering that a model trained in this manner only recognizes the enhanced images and loses the ability to distinguish original images, this paper incorporates both original and enhanced images into the classification experiments. In other words, the new learning objective is to minimize the following target based on the parameters of M :

$$\sum_{i=1}^n \lambda \text{loss}(M(x_i), y_i) + \varphi \text{loss}(M(x_i^E), y_i) \quad (2)$$

where λ and φ control the importance of training loss for original and enhanced images. When $\lambda = 1$ and $\varphi = 0$, the objective function in Eq. (2) simplifies to Eq. (1). In this paper, both λ and φ are set to 1, considering the importance of original and enhanced images to be of equal weight. Cross-entropy loss is employed to construct the loss functions in Eq. (1) and Eq. (2).

Considering that in real clinical scenarios, doctors need to classify and identify original images, all test set images in this experiment are unenhanced images. The model testing can be expressed as:

$$\hat{y} = f(M(x)) \quad (3)$$

where f is an output activation function (e.g., a sigmoid function, or a softmax function).

3.3. Classification model performance evaluation with SAM-IE

In this medical image enhancement and classification comparative experiment, we utilize the ResNet50 model and Swin Transformer model as the cornerstone of our methodology, given their effectiveness in medical image classification tasks and handling complex visual data. Firstly, we curated a comprehensive dataset comprising medical images related to four different disease classification tasks across various image modalities, ensuring accurate labeling and annotation. Subsequently, the dataset is partitioned into training and testing subsets to facilitate model evaluation. Next, we preprocess the images by resizing them to standard dimensions and applying normalization techniques to ensure dataset consistency. Furthermore, data augmentation techniques such as rotation and flipping are employed to enhance dataset robustness. For model architecture, we load pre-trained ResNet50 model and Swin Transformer model, leveraging their learned features to expedite training and enhance performance. During training, we fine-tune the two classification models on the medical image dataset, adjusting the final fully connected layers to accommodate the specific classification tasks. We monitor model performance using accuracy and F1 score metrics. Finally, we evaluate the model's generalization ability by assessing its performance on the test dataset and compare the classification results of various medical images before and after SAM-IE enhancement.

4. Experiments

4.1. Datasets

In order to demonstrate the effectiveness of SAM-IE in improving the medical image classification performance, this paper conducted experiments on four publicly available datasets: breast ultrasound image (BUSI) dataset (Al-Dhabyani, Gomaa, Khaled, & Fahmy, 2020), Massachusetts General Hospital (MGH) breast dataset (Dong et al., 2014), Human Against Machine with 10,000 training images (HAM10000) dataset (Philipp Tschandl, Rosendahl, & Kittler, 2018), and Fundus Multi-disease dataset (Pachade et al., 2021).

The BUSI dataset is a commonly used ultrasound dataset for breast tumors collected in 2018. The number of patients was 600 women. The dataset consists of 780 images with an average image size of 500×500 pixels. The image is in PNG format. The images were classified as normal, benign or malignant. In this paper, benign and malignant images were selected to be included in the classification experiment. The MGH Breast dataset is a binary classification pathology image dataset approved by the Partners Human Research Committee (Partners IRB). In this study, 233 images of ductal carcinoma in situ (DCIS) and 110 images of usual ductal hyperplasia (UDH) from MGH are included. The HAM10000 dataset contains a large number of multi-source

dermatoscopic images of common pigmented skin lesions (Philipp Tschandl et al., 2018). The HAM10000 dataset is a multi-class dataset, and this paper selected melanoma (MEL) and benign keratosis (BKL) for classification experiments. MEL is a malignant neoplasm derived from melanocytes that may appear in different variants. If excised in an early stage it can be cured by simple surgical excision. Melanomas can be invasive or noninvasive (in situ) (Schiffner et al., 2000; P. Tschandl, Rosendahl, & Kittler, 2015). BKL is a generic class that includes seborrheic keratoses, solar lentigo and lichen-planus like keratoses (Zaballos et al., 2010). The Fundus Multi-disease is a fundus images dataset consisting of 3200 images along with the expert annotations divided into two categories: normal and abnormal. The detailed structure of the datasets is shown in Table 1.

4.2. Implementation details and evaluation metrics

We implement all models using the PyTorch framework on Ubuntu 20.04 system, which is equipped with an Intel(R) Xeon(R) Platinum 8269CY CPU @ 2.50 GHz, alongside the NVIDIA GeForce RTX 3090 24 GB.

This paper selects the ResNet50 and Swin Transformer as the classification models for medical image classification. For ResNet50, the initial learning rate is set to 1×10^{-3} , and the batch size is set to 80. Experiments are conducted on the four datasets both before and after SAM-IE, and the classification performance is compared. As for the Swin Transformer, the initial learning rate is set to 1×10^{-4} , and the batch size is set to 48. Similarly, experiments are conducted on the four datasets before and after SAM-IE, and the classification performance is compared.

The classification performance before and after SAM-IE with different classification methods is evaluated using commonly used metrics in medical image classification: the area under the receiver operating characteristic curve (AUC), accuracy, precision, sensitivity, specificity, Youden's index (YI) and F1 score. Additionally, to highlight the differences in classification performance before and after SAM-IE, this paper employs the Delong test to examine the differences in AUC values between the results before and after SAM enhancement.

5. Results

With the support of SAM-IE, this study accomplished image enhancement for the BUSI, MGH Breast, HAM10000, and Fundus Multi-disease datasets, as depicted in Fig. 2. The initial row exhibits examples of unaltered images from the four datasets. The second and third rows showcase the binary masks and contour masks generated by SAM for each example. The enhanced images by SAM-IE are presented in the last row of Fig. 2. To highlight the effectiveness of SAM-IE on medical image classification models, this paper compares the classification performance on two networks (ResNet50 and Swin Transformer) before and after using SAM-IE. Moreover, we have also supplemented the classification performance of the four datasets used in this paper on seven other common classification models, including DenseNet121, DenseNet161, DenseNet169, ResNet18, ResNet34, ResNeXt50_32 $\times$ 4D, and

Table 1

Statistical description of the BUSI, MGH Breast, HAM10000, and Fundus Multi-disease datasets.

Dataset	Label	Overall	Training	Testing
BUSI	Benign	437	350	87
	Malignant	210	168	42
MGH Breast	UDH	110	88	22
	DCIS	233	186	47
HAM10000	MEL	1284	1113	171
	BKL	1316	1099	217
Fundus Multi-disease	Normal	669	535	134
	Abnormal	2531	2025	506

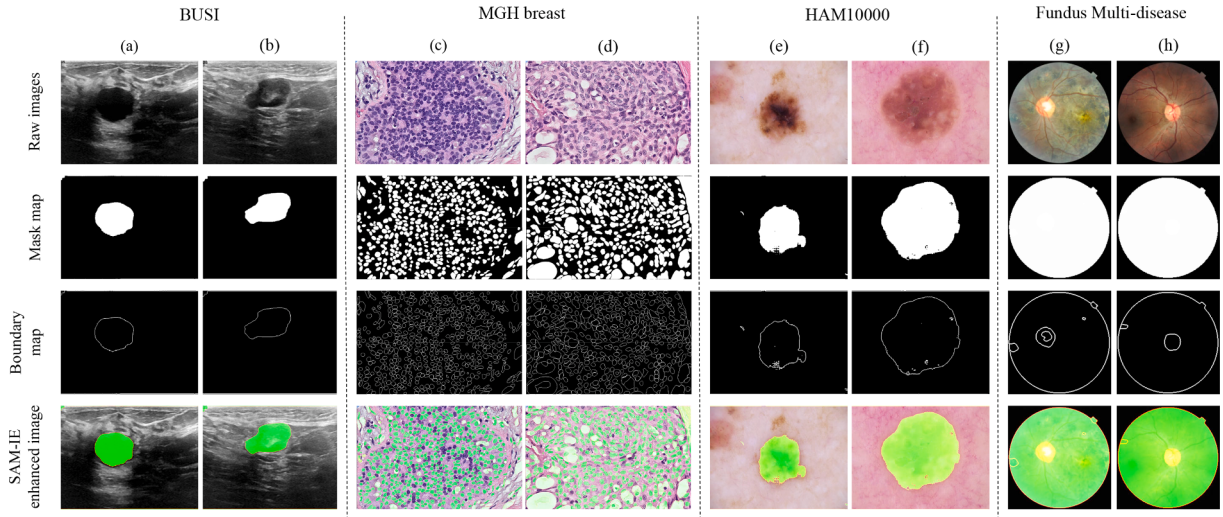


Fig. 2. Illustration of image enhancement with SAM-based Image Enhancement (SAM-IE). Two typical examples from each dataset were selected.

ResNeXt101_32 $\times$ 8D. The classification results on the four datasets are shown in Table 2 and Table 3. Furthermore, to visually illustrate the differences in classification results between with and without SAM-IE for each set of experiments, receiver operating characteristic (ROC) curves were plotted for each set of results, as depicted in Fig. 3.

As shown in Fig. 4, we use Grad-CAM++ to generate class activation heatmaps of the test set images based on ResNet50 for visualizing the contribution distribution of different regions in the input images to the predicted outputs. In the heatmaps, areas with deeper red colors indicate higher values, indicating that the corresponding regions have higher response and contribution to the network. Thus, we can compare the differences in contribution distribution in medical images before and after using SAM-IE in the classification network through the heatmaps.

5.1. Breast tumor classification on BUSI dataset

Fig. 2a and Fig. 2b display ultrasound images of benign and malignant breast tumors, showcasing the image enhancement processes with SAM-IE. The SAM-IE enhancement method effectively highlights tumor

regions and boundaries in the ultrasound images. The classification experiments were conducted using the ResNet50 and the Swin Transformer models with and without SAM-IE. From Table 2, it can be seen that the classification results of the BUSI images enhanced by SAM-IE on both the ResNet50 and Swin Transformer models have shown significant improvements over the original images. All classification indices demonstrate noteworthy improvement. Furthermore, the classification results of the enhanced BUSI images on the ResNet50 model are superior to those on the Swin Transformer model. Compared to the classification results of other models, the enhanced BUSI images still demonstrate better performance on the ResNet50 model.

ROC curves with and without SAM-IE for both ResNet50 and Swin Transformer models are depicted in Fig. 3a. For the ResNet50 model, the AUC increased from 0.943 to 0.981 with SAM-IE enhancement (Delong test: $Z = -2.59$, $p < 0.05$). For the Swin Transformer model, the AUC increased from 0.927 to 0.975 with SAM-IE enhancement (Delong test: $Z = -2.13$, $p < 0.05$). From Fig. 4, it can be seen that after enhancement by SAM-IE, the tumor regions in the BUSI images contribute more to the ResNet50 classification network, aiding in the accurate determination of

Table 2

Classification results with our proposed SAM-IE on the BUSI and MGH Breast datasets based on ResNet50 and Swin Transformer (bottom two rows for each dataset), compared with results of classic classification networks without SAM-IE including the DenseNet121, DenseNet161, DenseNet169, ResNet18, ResNet34, ResNet50, ResNeXt50_32 $\times$ 4D, ResNeXt101_32 $\times$ 8D, and Swin Transformer.

Dataset	Model	AUC	Accuracy	Precision	Sensitivity	Specificity	YI	F1 score
BUSI	DenseNet121	0.925	0.853	0.717	0.905	0.828	0.732	0.800
	DenseNet161	0.919	0.868	0.755	0.881	0.862	0.743	0.813
	DenseNet169	0.898	0.868	0.755	0.881	0.862	0.743	0.813
	ResNet18	0.915	0.892	0.850	0.810	0.931	0.741	0.829
	ResNet34	0.910	0.884	0.787	0.881	0.885	0.766	0.832
	ResNet50	0.943	0.884	0.776	0.905	0.874	0.778	0.835
	ResNeXt50_32 $\times$ 4D	0.939	0.923	0.848	0.929	0.920	0.848	0.886
	ResNeXt101_32 $\times$ 8D	0.952	0.930	0.902	0.881	0.954	0.835	0.892
	Swin Transformer	0.927	0.876	0.771	0.881	0.874	0.755	0.822
	ResNet50 w/ SAM-IE	0.981	0.946	0.889	0.952	0.943	0.895	0.920
	Swin Transformer w/ SAM-IE	0.975	0.961	0.974	0.905	0.989	0.893	0.938
MGH Breast	DenseNet121	0.860	0.783	0.971	0.702	0.955	0.657	0.815
	DenseNet161	0.894	0.826	1.000	0.745	1.000	0.745	0.854
	DenseNet169	0.902	0.855	1.000	0.787	1.000	0.787	0.881
	ResNet18	0.873	0.812	0.886	0.830	0.773	0.603	0.857
	ResNet34	0.887	0.841	0.929	0.830	0.864	0.693	0.876
	ResNet50	0.883	0.870	1.000	0.809	1.000	0.809	0.894
	ResNeXt50_32 $\times$ 4D	0.906	0.913	0.902	0.979	0.773	0.752	0.939
	ResNeXt101_32 $\times$ 8D	0.912	0.826	0.973	0.766	0.955	0.721	0.857
	Swin Transformer	0.893	0.826	0.973	0.766	0.955	0.721	0.857
	ResNet50 w/SAM-IE	0.981	0.928	0.977	0.915	0.955	0.869	0.945
	Swin Transformer w/ SAM-IE	0.980	0.942	0.939	0.979	0.864	0.842	0.958

Table 3

Classification results with our proposed SAM-IE on the HAM10000 and Fundus Multi-disease datasets based on ResNet50 and Swin Transformer (bottom two rows for each dataset), compared with results of classic classification networks without SAM-IE including the DenseNet121, DenseNet161, DenseNet169, ResNet18, ResNet34, ResNet50, ResNeXt50_32 × 4D, ResNeXt101_32 × 8D, and Swin Transformer.

Dataset	Model	AUC	Accuracy	Precision	Sensitivity	Specificity	YI	F1 score
HAM10000	DenseNet121	0.852	0.778	0.703	0.860	0.714	0.574	0.774
	DenseNet161	0.847	0.802	0.783	0.760	0.834	0.594	0.772
	DenseNet169	0.841	0.789	0.731	0.825	0.760	0.585	0.775
	ResNet18	0.908	0.851	0.805	0.871	0.834	0.705	0.837
	ResNet34	0.862	0.804	0.741	0.854	0.765	0.619	0.794
	ResNet50	0.906	0.851	0.816	0.854	0.848	0.702	0.834
	ResNeXt50_32 × 4D	0.926	0.879	0.883	0.836	0.912	0.749	0.859
	ResNeXt101_32 × 8D	0.912	0.845	0.776	0.912	0.793	0.705	0.839
	Swin Transformer	0.902	0.820	0.742	0.906	0.751	0.658	0.816
	ResNet50 w/ SAM-IE	0.931	0.879	0.869	0.854	0.899	0.752	0.861
	Swin Transformer w/ SAM-IE	0.937	0.869	0.797	0.942	0.811	0.753	0.863
Fundus Multi-disease	DenseNet121	0.934	0.831	0.995	0.791	0.985	0.776	0.881
	DenseNet161	0.932	0.838	0.981	0.810	0.940	0.751	0.887
	DenseNet169	0.933	0.842	0.981	0.816	0.940	0.757	0.891
	ResNet18	0.942	0.888	0.982	0.874	0.940	0.814	0.925
	ResNet34	0.951	0.863	0.998	0.828	0.993	0.821	0.905
	ResNet50	0.946	0.883	0.970	0.879	0.896	0.775	0.922
	ResNeXt50_32 × 4D	0.948	0.855	0.988	0.826	0.963	0.789	0.900
	ResNeXt101_32 × 8D	0.946	0.870	0.980	0.854	0.933	0.787	0.912
	Swin Transformer	0.946	0.880	0.976	0.870	0.918	0.788	0.920
	ResNet50 w/SAM-IE	0.955	0.889	0.968	0.889	0.888	0.777	0.927
	Swin Transformer w/ SAM-IE	0.959	0.903	0.985	0.891	0.948	0.839	0.936

the benignity or malignancy of breast tumors. In contrast, with the original images, the classification network often overly focuses on artifacts in the ultrasound images, which can easily lead to diagnostic errors.

5.2. Breast disease classification on MGH Breast dataset

Fig. 2c and Fig. 2d respectively depict two types of pathological images, UDH and DCIS, along with their enhancement processes with SAM-IE. The SAM-IE enhancement method emphasizes the characteristics of cell regions and boundaries in the pathological images. Using the ResNet50 model and Swin Transformer model, this study conducts classification experiments with and without SAM-IE. Table 2 shows that compared with the original images, the MGH Breast images enhanced by SAM-IE achieved better AUC and accuracy on both the ResNet50 and Swin Transformer models. The results of SAM-IE also showed more balanced sensitivity and specificity, with significant improvements in YI and F1 scores. Moreover, the classification results of the MGH Breast enhanced images on the ResNet50 model were superior to those on the Swin Transformer model. Compared with the classification results of other models, the MGH Breast enhanced images still exhibited superior performance on the ResNet50 model.

ROC curves with and without SAM-IE on the ResNet50 and Swin Transformer models are illustrated in Fig. 3b. Incorporating SAM-IE model led to a significant improvement in AUC for both the ResNet50 and Swin Transformer models. Specifically, for the ResNet50 model, the AUC increased from 0.883 to 0.981 (Delong test: $Z = -2.44$, $p < 0.05$). For the Swin Transformer model, the AUC increased from 0.893 to 0.980 with SAM-IE enhancement (Delong test: $Z = -2.33$, $p < 0.05$). From Fig. 4, it is evident that after enhancement by SAM-IE, the effective cellular areas in the MGH Breast images contribute more significantly to the ResNet50 classification network. In contrast, with the original images, the classification network is unable to adequately focus on the pathological areas that aid in diagnosis.

5.3. Skin lesions classification on HAM10000 dataset

Fig. 2e and Fig. 2f illustrate the dermatoscopic images of MEL and BKL in the HAM10000 dataset alongside their enhancement processes with SAM-IE. The SAM-IE model implemented in this study notably

accentuates the skin lesion area and sharpens the contour between the lesion tissue and the surrounding tissue in the dermatoscopic images. Using both ResNet50 and Swin Transformer models, this study conducted classification experiments with and without SAM-IE. Table 3 shows that compared to the original images, the images enhanced by SAM-IE demonstrate superior classification performance on both the ResNet50 and Swin Transformer models, with significant improvements in accuracy, precision, and specificity. The enhancement effect of SAM-IE significantly improves the classification performance of BKL images. Moreover, the classification results of the HAM10000 enhanced images on the Swin Transformer model are superior to those on the ResNet50 model. Compared to the classification results of other models, the HAM10000 enhanced images still exhibit superior performance on the Swin Transformer model.

The ROC curves, depicted in Fig. 3c, further highlight the positive impact of enhancement with SAM-IE on classification performance for both models. For the ResNet50 model, the AUC increased from 0.906 to 0.931 with SAM-IE model (Delong test: $Z = -2.13$, $p < 0.05$). For the Swin Transformer model, the AUC witnessed an improvement from 0.902 to 0.937 with SAM-IE model (Delong test: $Z = -2.65$, $p < 0.05$). From Fig. 4, it can be seen that after enhancement with SAM-IE, the skin cancer lesion areas in the HAM10000 images contribute more significantly to the ResNet50 classification network, aiding in the accurate determination of the cancer type. In contrast, for the original images, the classification network is unable to effectively focus on the lesion areas.

5.4. Fundus state classification on fundus Multi-disease dataset

Fig. 2g and Fig. 2h display fundus images under normal and abnormal conditions, respectively, along with the enhancement process by SAM-IE model. In the binary mask, it is evident that SAM enhances the visual field of the fundus image. However, the effectiveness of SAM-IE model in distinguishing optic disc, optic cup, and background areas is not prominent in this context. Conversely, the contour mask illustrates a clearer delineation of the optic disc boundary, but SAM-IE still exhibits limitations in identifying the optic cup.

Employing both ResNet50 and Swin Transformer models, this paper conducts classification experiments with and without SAM-IE model. Table 3 reveals that, in comparison to the original image, enhanced results by SAM-IE on the ResNet50 model exhibit slight improvements,

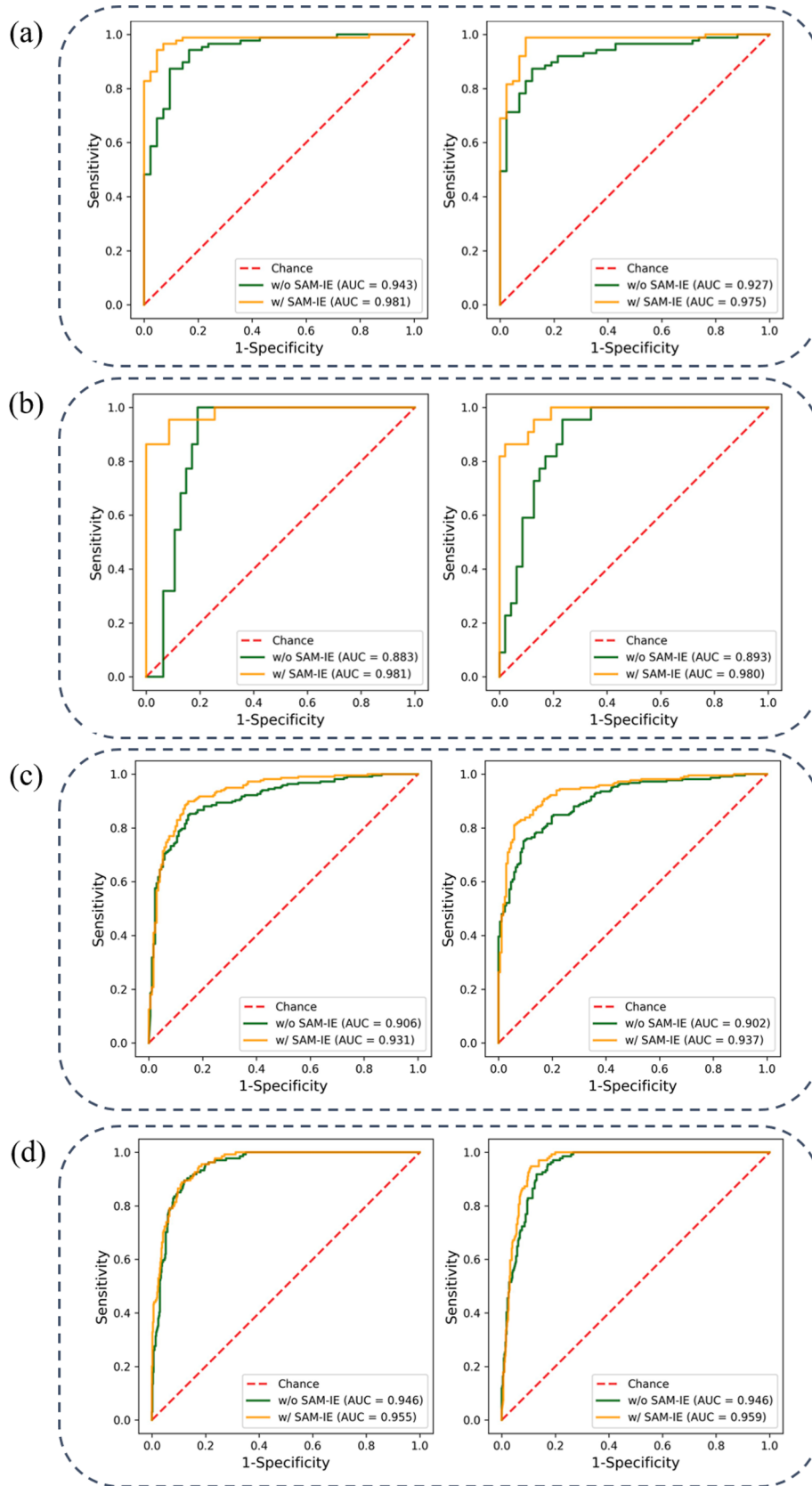


Fig. 3. Receiver operating characteristic (ROC) curves for classification with and without SAM-IE based on ResNet50 (left) and Swin Transformer (right) models on the BUSI (a), MGH Breast (b), HAM10000 (c), and Fundus Multi-disease (d) datasets.

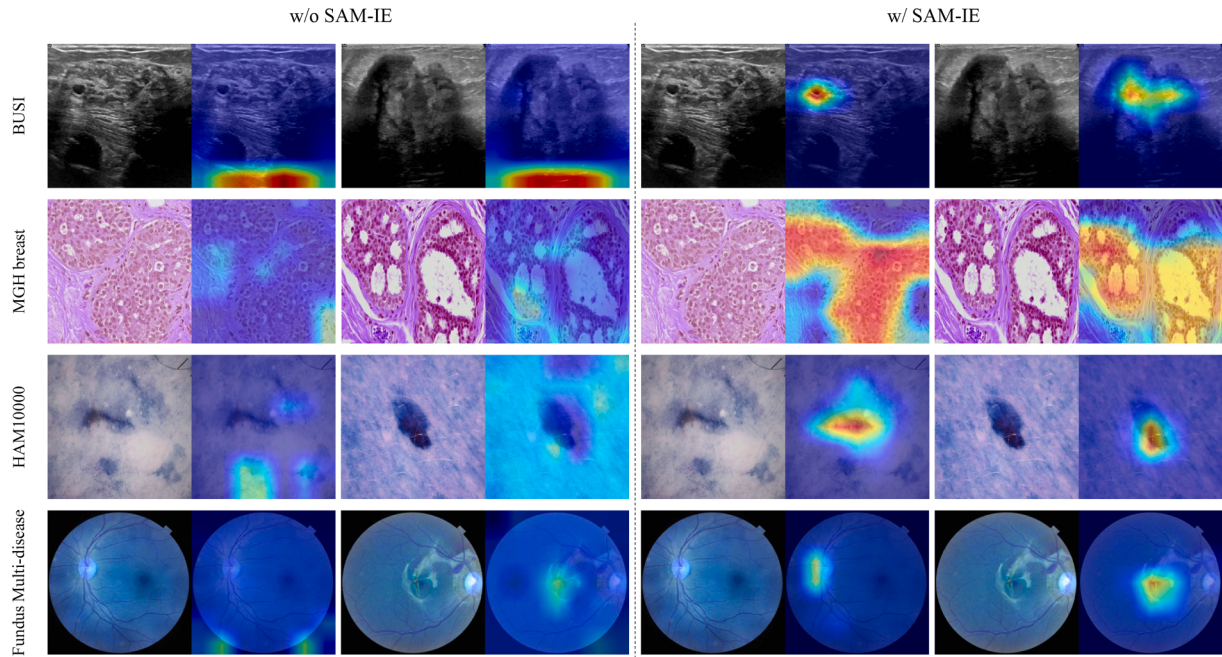


Fig. 4. The class activation heatmaps generated from the test set images based on the ResNet50 model. Two typical examples from each dataset were selected. The heatmaps on the left show the results without SAM-IE, while the heatmaps on the right show the results with SAM-IE.

with significant enhancements in sensitivity. These results indicate that fundus image enhancement by SAM-IE can improve the classification performance of abnormal images to some extent. For the Swin Transformer model, compared to the original image, enhanced fundus images achieve superior classification results, with notable improvements in various classification indicators. Additionally, the classification results of enhanced fundus images are superior on the Swin Transformer model compared to the ResNet50 model. In comparison to the results of other models, the classification outcomes of enhanced fundus images still demonstrate superiority on the Swin Transformer model.

The ROC curves, depicted in Fig. 3d, provide a visual representation of the impact of fundus image enhancement by SAM-IE on classification performance for both ResNet50 and Swin Transformer models. For the ResNet50 model, the AUC witnessed a marginal increase from 0.946 to 0.955 with SAM-IE model (Delong test: $Z = -1.50, p = 0.13$). Meanwhile, for the Swin Transformer model, the AUC increased from 0.946 to 0.959 with SAM-IE model (Delong test: $Z = -2.10, p < 0.05$). From Fig. 4, it can be observed that after SAM-IE enhancement, the pathological regions in the fundus images contribute more to the ResNet50 classification network, aiding in accurately determining the abnormalities in the fundus images. In contrast, the classification network's ability to focus on the pathological regions is poorer in the original images.

6. Discussion

In recent years, the advent and rapid advancement of SAM have yielded commendable achievements in image segmentation, particularly in natural image segmentation, where SAM can achieve more accurate results when guided by prompts. However, due to the substantial differences between medical and natural images, SAM's segmentation performance in medical images falls short of satisfactory levels. The effective application of SAM in medical image processing and disease diagnosis remains extensive attention. During exploration, it was observed that while SAM may not precisely segment lesion or tissue regions, it could output the most probable region of interest in an image by limiting the confidence level of the segmentation region. In clinical diagnosis, radiologists typically focus on morphological and structural changes in lesion or tissue areas and their surroundings. Hence, in this

study, SAM is employed to extract regions of interest with certain probabilities in medical images. These regions and their boundaries are then highlighted and emphasized in the original image to generate enhanced images. To assess the potential impact of enhanced images by SAM-IE on disease diagnosis, comparative experiments are designed with and without SAM-IE, using the ResNet50 model and Swin Transformer model, respectively.

According to the classification results, all performance metrics for both the ResNet50 and Swin Transformer models, using the BUSI and HAM10000 dataset after enhancement by SAM-IE, have displayed significant improvements, with discernible statistical difference in the AUC (Table 2, Table 3, Fig. 3a and Fig. 3c). By observing Fig. 2, we can find that the images in these datasets contain a single target region, and enhancement operation with SAM-IE accurately identifies and locates the region of interest within the image. Furthermore, considering the sizable lesion area and distinct lesion morphology in these images, the enhancement with SAM-IE demonstrates advantageous in the diagnostic process. For the enhanced images of MGH Breast dataset, it's noticeable that cells in the images are not entirely recognized. However, when comparing Fig. 2c and Fig. 2d, there are fewer unrecognized cells in Fig. 2d, indicating a better enhancement effect on DCIS images. Consequently, the sensitivity in the classification results with SAM-IE on the MGH Breast dataset surpasses that without SAM-IE (Table 2). Due to the inherent characteristics of fundus images, the SAM-IE can enhance the entire circular visual field of fundus images without specific prompts. Although the area and boundaries of optic disc and optic cup are highlighted to some extent, the contour between the optic cup and the surrounding area remains unclear in the enhanced image (Fig. 2). The classification results of SAM-IE enhanced fundus images demonstrated improvement on both ResNet50 model and Swin Transformer model compared to those without SAM-IE enhancement. However, this improvement effect is not significant (Table 3). On the other side, the comparison of ROC curves on the ResNet50 model in Fig. 3d indicate no significant difference, which also shows the improvement effect is weak. Moreover, it can be observed that compared to the experimental results of other 7 classification models without SAM-IE, the classification results after using SAM-IE image enhancement in this paper still exhibit superiority (Table 2, Table 3). From Fig. 4, it can be seen that after SAM-IE

enhancement, the pathological regions in the images can contribute more to the classification network, thereby improving the accuracy of classification. In contrast, for the original images, the classification network tends to overly focus on the distracting factors in the images, leading to diagnostic errors.

Based on the aforementioned experimental results, it is evident that the SAM-IE model effectively accentuates lesions or tissue areas in the images, consequently improving the classification accuracy of medical image. Notably, the SAM-IE model demonstrates superior effects on images featuring a single object, large area, and clear boundaries. However, it is susceptible to interference from additional information in the image, compromising the enhancement of the target region.

To further improve performance of SAM-IE in medical image classification, several feasible strategies can be considered. Firstly, the incorporation of a prompt module into the SAM-IE could enhance the accuracy of recognizing the enhanced region in the images. Secondly, the impact of SAM-IE can be assessed using various medical image classification models to bolster the reliability of the SAM-IE model. Additionally, a comprehensive evaluation of the SAM-IE model's impact on medical image classification could be conducted through a collaborative approach that involves both radiologists and classification models.

7. Conclusions

The classification results across the four public datasets demonstrate that SAM-IE contributes to an improvement in the performance of medical image classification models. Moreover, the SAM-IE model proposed in this study exhibits particular suitability for images characterized by a large area, a small number of target areas, and significant morphological differences from the surrounding regions. By accentuating the target region and its contour in the image, the SAM-IE model elevates the classification performance of the ResNet50 and Swin Transformer models. This verification process not only emphasizes the efficacy of SAM-IE for disease diagnosis on various imaging modalities but also expands the application scope of SAM in medical image analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [62071285]; and the Shanghai Technical Service Center of Science and Engineering Computing, Shanghai University.

References

- Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief*, 28, Article 104863. <https://doi.org/10.1016/j.dib.2019.104863>
- Chai, S., Jain, R. K., Teng, S., Liu, J., Li, Y., Tateyama, T., & Yen-wei, C. (2023). Ladder Fine-tuning approach for SAM integrating complementary network. *arXiv.org*.doi:10.48550/arxiv.2306.12737.
- Cui, C., Deng, R., Liu, Q., Yao, T., Bao, S., Remedios, L. W., Tang, Y., & Huo, Y. (2023). All-in-SAM: from Weak Annotation to Pixel-wise Nuclei Segmentation with Prompt-based Finetuning. *arXiv.org*.doi:10.48550/arxiv.2307.00290.
- Dai, H., Ma, C., Liu, Z., Li, Y., Peng, S., Xiaozheng, W., Zhao, L., Wu, Z., Zeng, F., Zhu, D., Liu, W., Li, Q., Liu, T., & Li, X. (2023). SAMAug: Point Prompt Augmentation for Segment Anything Model. *arXiv.org*.doi:10.48550/arxiv.2307.01187.
- de Araujo, A. F., Constantinou, C. E., & Tavares, J. M. R. S. (2014). New artificial life model for image enhancement. *Expert Systems with Applications*, 41, 5892–5906. <https://doi.org/10.1016/j.eswa.2014.03.029>
- Deng, G., Zou, K., Ren, K., Wang, M., Yuan, X., Sancong, Y., & Fu, H. (2023). SAM-U: Multi-box prompts triggered uncertainty estimation for reliable SAM in medical image. *arXiv.org*.doi:10.48550/arxiv.2307.04973.
- Deng, R., Cui, C., Liu, Q., Yao, T., Remedios, L. W., Bao, S., Landman, B. A., Wheless, L. E., Coburn, L. A., Wilson, K. T., Wang, Y., Zhao, S., Fogo, A. B., Yang, H., Tang, Y., & Huo, Y. (2023). Segment Anything Model (SAM) for Digital Pathology: Assess Zero-shot Segmentation on Whole Slide Imaging. *arXiv.org*.doi:10.48550/arxiv.2304.04155.
- Dinh, P. H., & Giang, N. L. (2022). A new medical image enhancement algorithm using adaptive parameters. *International Journal of Imaging Systems and Technology*, 32, 2198–2218. <https://doi.org/10.1002/ima.22778>
- Dong, F. I. H., Oh, E.-Y., Lerwill, M. F., Brachtel, E. F., Jones, N. C., et al. (2014). Computational pathology to discriminate benign from malignant intraductal proliferations of the breast. *PLoS ONE*, 9(12). <https://doi.org/10.1371/journal.pone.0114885>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houshy, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv.org*.doi:10.48550/arxiv.2010.11929.
- Edward J. Hu, Y. S., Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen. (2021). LoRA: Low-Rank Adaptation of Large Language Models. *arXiv preprint*, arXiv:2106.09685.
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., & Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, 30, 1323–1341. <https://doi.org/10.1016/j.mri.2012.05.001>
- Gao, Y., Xia, W., Hu, D., & Gao, X. (2023). DeSAM: Decoupling Segment Anything Model for Generalizable Medical Image Segmentation. *arXiv.org*.doi:10.48550/arxiv.2306.00499.
- Hu, C., Xia, T., Ju, S., & Li, X. (2023). When SAM Meets Medical Images: An Investigation of Segment Anything Model (SAM) on Multi-phase Liver Tumor Segmentation. *arXiv.org*.doi:10.48550/arxiv.2304.08506.
- Hu, M., Li, Y., & Yang, X. (2023). SkinSAM: Empowering Skin Cancer Segmentation with Segment Anything Model. *arXiv.org*.doi:10.48550/arxiv.2304.13973.
- Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., Chi, H., Hu, X., Deng-Ping, F., Dong, F., & Ni, D. (2023). Segment Anything Model for Medical Images? *arXiv.org*.doi:10.48550/arxiv.2304.14660.
- Ji, G.-P., Fan, D.-P., Xu, P., Zhou, B., Cheng, M.-M., & Van Gool, L. (2023). SAM struggles in concealed scenes — empirical study on “Segment Anything”. *Science China Information Sciences*, 66. <https://doi.org/10.1007/s11432-023-3881-x>
- Ji, W., Li, J., Bi, Q., Liu, T., Li, W., & Cheng, L. (2023). Segment Anything Is Not Always Perfect: An Investigation of SAM on Different Real-world Applications. *arXiv.org*.doi:10.48550/arxiv.2304.05750.
- Kaiming, H., Xiangyu, Z., Shaoqing, R., & Jian, S. Deep Residual Learning for Image Recognition. In (pp. 770–778): IEEE.
- Kirillov, A., Mintun, E., Nikhila, R., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Wan-Yen, L., Dollár, P., & Girshick, R. (2023). Segment Anything. *arXiv.org*.doi:10.48550/arxiv.2304.02643.
- Lei, W., Xu, W., Zhang, X., Kang, L., & Zhang, S. (2023). MedLSAM: Localize and Segment Anything Model for 3D CT Images. *arXiv.org*.doi:10.48550/arxiv.2306.14752.
- Li, Y., Hu, M., & Yang, X. (2023). Polyp-SAM: Transfer SAM for Polyp Segmentation. *arXiv.org*.doi:10.48550/arxiv.2305.00293.
- Liu, Y., Zhang, J., She, Z., Kheradmand, A., & Mehran, A. (2023). SAMM (Segment Any Medical Model): A 3D Slicer Integration to SAM. *arXiv.org*.doi:10.48550/arxiv.2304.05622.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In (pp. 9992–10002). Piscataway: Piscataway: IEEE.
- Ma, J., He, Y., Li, F., Lin, H., You, C., & Wang, B. (2023). Segment Anything in Medical Images. *arXiv.org*.doi:10.48550/arxiv.2304.12306.
- Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., & Zhang, Y. (2023). Segment anything model for medical image analysis: An experimental study. *Medical Image Analysis*, 89, 102918. <https://doi.org/10.1016/j.media.2023.102918>
- Ortiz, A., Górriz, J. M., Ramírez, J., Salas-González, D., & Llamas-Elvira, J. M. (2013). Two fully-unsupervised methods for MR brain image segmentation using SOM-based strategies. *Applied Soft Computing*, 13, 2668–2682. <https://doi.org/10.1016/j.asoc.2012.11.020>
- Pachade, S., Porwal, P., Thulkar, D., Kokare, M., Deshmukh, G., Sahasrabudhe, V., Giancardo, L., Queller, G., & Mériaudeau, F. (2021). Retinal Fundus Multi-Disease Image Dataset (RFMiD): A Dataset for Multi-Disease Detection Research. *Data (Basel)*, 6, 14. <https://doi.org/10.3390/data6020014>
- Ronneberger, O., Fischer, P., & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In (pp. 234–241). Cham: Cham: Springer International Publishing.
- Rundo, L., Tangherloni, A., Nobile, M. S., Militello, C., Besozzi, D., Mauri, G., & Cazzaniga, P. (2019). MedGA: A novel evolutionary method for image enhancement in medical imaging systems. *Expert Systems with Applications*, 119, 387–399. <https://doi.org/10.1016/j.eswa.2018.11.013>
- Schiffner, R., Schiffner-Rohe, J., Vogt, T., Landthaler, M., Wlotzke, U., Cognetta, A. B., & Stolz, W. (2000). Improvement of early recognition of lentigo maligna using

- dermatoscopy. *Journal of the American Academy of Dermatology*, 42, 25–32. [https://doi.org/10.1016/S0190-9622\(00\)90005-7](https://doi.org/10.1016/S0190-9622(00)90005-7)
- Shaharabany, T., Dahan, A., Giryas, R., & Wolf, L. (2023). AutoSAM: Adapting SAM to Medical Images by Overloading the Prompt Encoder. *arXiv.org*.doi:10.48550/arxiv.2306.06370.
- Tschanl, P., Rosendahl, C., & Kittler, H. (2015). Dermatoscopy of flat pigmented facial lesions. *Journal of the European Academy of Dermatology and Venereology*, 29, 120–127. <https://doi.org/10.1111/jdv.12483>
- Tschanl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, Article 180161. <https://doi.org/10.1038/sdata.2018.161>
- Wu, J., Zhang, Y., Rao, F., Fang, H., Liu, Y., Wang, Z., Xu, Y., & Jin, Y. (2023). Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation. *arXiv.org*.doi:10.48550/arxiv.2304.12620.
- Zaballos, P., Salsench, E., Serrano, P., Cuellar, F., Puig, S., & Malveyh, J. (2010). Studying regression of seborrheic keratosis in lichenoid keratosis with sequential dermoscopy imaging. *Dermatology*, 220, 103–109. <https://doi.org/10.1159/000265556>
- Zhang, K., & Liu, D. (2023). Customized Segment Anything Model for Medical Image Segmentation. *arXiv.org*.doi:10.48550/arxiv.2304.13785.
- Zhang, Y., & Jiao, R. (2023). Towards Segment Anything Model (SAM) for Medical Image Segmentation: A Survey. *arXiv.org*.doi:10.48550/arxiv.2305.03678.
- Zhang, Y., Zhou, T., Wang, S., Liang, P., & Chen, D. Z. (2023). Input Augmentation with SAM: Boosting Medical Image Segmentation with Segmentation Foundation Model. *arXiv.org*.doi:10.48550/arxiv.2304.11332.
- Zhou, T., Zhang, Y., Zhou, Y., Wu, Y., & Chen, G. (2023). Can SAM Segment Polyps? *arXiv.org*.doi:10.48550/arxiv.2304.07583.